Multi-user information theory 2

Lecture 6

Lecturer: Haim Permuter *Scribe: Ohad Elishco*

I. Capacity region of degraded broadcast channel

The broadcast channel is a communication channel in which there is one sender and several receivers, as presented in Fig. 1. The broadcast channel (BC) was first introduced by Cover [4].

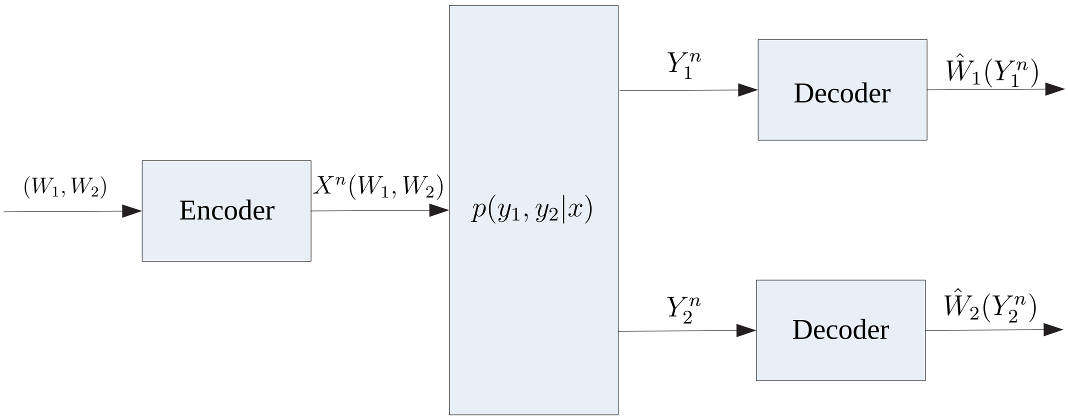

Fig. 1. Broadcast channel.

We begin with some basic definitions for the broadcast channel.

A. Definitions for Broadcast Channel

Definition 1 (Code for the BC) A *broadcast channel* (BC) consists of an input alphabet $\mathcal{X}$ and two output alphabets, $\mathcal{Y}_1$ and $\mathcal{Y}_2$, and a probability transition function $p(y_1, y_2|x)$. The channel called *memoryless* if

$$p(y_{1,i}, y_{2,i}|x^i, y_1^{i-1}, y_2^{i-1}) = p(y_{1,i}, y_{2,i}|x_i). \quad (1)$$

Next we define the code for the BC and the average probability of error.

Definition 2 A $((2^{nR_1}, 2^{nR_2}), n)$ code for the BC consists of two sets of integers $\mathcal{W}_1 = \{1, 2, \ldots, 2^{nR_1}\}$ and $\mathcal{W}_2 = \{1, 2, \ldots, 2^{nR_2}\}$, called the message sets. There is one encoding function

$$f : \mathcal{W}_1 \times \mathcal{W}_2 \to \mathcal{X}^n,$$

and two decoding functions:

$$g_1 : \mathcal{Y}_1^n \to \mathcal{W}_1$$

$$g_1 : \mathcal{Y}_2^n \to \mathcal{W}_2.$$

Definition 3 (Average probability of error) We define the average probability of error as the probability that the decoded messages are not equal to the transmitted message. That is, $P_e^{(n)} = \Pr\Big(\{W_1 \neq \hat{W}_1\} \cup \{W_2 \neq \hat{W}_2\}\Big) = \Pr\Big((W_1, W_2) \neq (\hat{W}_1, \hat{W}_2)\Big)$. We assume that (W_1, W_2) are distributed uniformly over $2^{nR_1} \times 2^{nR_2}$.

Now we define the achievable rate pair and the capacity region.

Definition 4 (Achievable rate pair) A rate pair (R_1, R_2) is said to be achievable for the BC if there exists a sequence of $((2^{nR_1}, 2^{nR_2}), n)$ codes with $P_e^{(n)} \to 0$ as $n \to \infty$.

Definition 5 (Capacity region) The capacity region is the closure of the union of all achievable rate pairs.

The general model has yet to be solved, therefore we discuss special cases of the BC. The case of physically degraded BC and the case of stochastically degraded BC.

B. Degraded Broadcast Channel

Definition 6 (Physically degraded BC) A broadcast channel is said to be *physically degraded* if

$$p(y_1, y_2|x) = p(y_1|x)p(y_2|y_1). \tag{2}$$

This definition means that we have the Markov chain $(Y_2 - Y_1 - X)$.

Definition 7 (Stochastically degraded BC) A broadcast channel is said to be *stochastically degraded* if there exists a distribution $p'(y_2|y_1)$ such that

$$p(y_2|x) = \sum_{y_1} p(y_1|x)p'(y_2|y_1). \tag{3}$$

Note that it is always the case: $p(y_2|x) = \sum_{y_1} p(y_1, y_2|x)$. Hence, the stochastically degraded BC means that there exists some $p'(y_2|y_1)$ such that

$$P'(y_1, y_2|x) = P(y_1|x)P'(y_2, y_1) \tag{4}$$

induces the same conditional probability $P(y_1|x)$ and $P(y_2|x)$ of the original channel $P(y_1, y_2|x)$. Namely, $P'(y_i|x) = P(y_i|x) \quad i = 1, 2$. We call $P'(y_1, y_2|x)$ given in (3) the associated phisically degraded BC.

We now show that the capacity region of the stochastically degraded BC is equal to the associate physically degraded BC.

Lemma 1 For a stochastically degraded BC, the capacity region of the channel defined with the probability $P'(y_1, y_2|x)$ (as in (4)) that satisfies (3) is equal to the capacity region of the stochastically degraded channel defined with the probability $P(y_1, y_2|x)$.

Proof: We denote A as the event $\{\hat{W}_1(Y_1^n) \neq w_1\}$ and $B = \{\hat{W}_2(Y_2^n) \neq w_2\}$. Recall that $P_e^n = \Pr\left(A \cup B | W_1 = w_1, W_2 = w_2\right)$. Hence, by fixing (w_1, w_2) we fix $x^n(w_1, w_2)$ and we have that the expression

$$\begin{aligned} \Pr(A|W_1 = w_1, W_2 = w_2) &= \Pr(A|W_1 = w_1, W_2 = w_2, x^n(w_1, w_2)) \\ &= \Pr(A|x^n) \end{aligned}$$

depends only on $P(y_1^n|x^n)$ which is equal to $\prod_{i=1}^n = P(y_i|x_i)$ due to the memoryless property. Similarily, we have that $\Pr(B)$ depends only on $P(y_2|x)$. By the upper bound $\Pr(A \cup B|x^n) \leq \Pr(A|x^n) + \Pr(B|x^n)$. Moreover, $\min\{\Pr(A|x^n), \Pr(B|x^n)\} \leq \Pr(A \cup B|x^n)$. Therefore, $\Pr(A \cup B|x^n)$ is going to 0 if and only if $\Pr(A|x^n)$ and $\Pr(B|x^n)$ are going to 0. We have that $P_e^{(n)} \leq \Pr\left(A|W_1 = w_1, W_2 = w_2, x^n(w_1, w_2)\right) + \Pr\left(B|W_1 = w_1, W_2 = w_2, x^n(w_1, w_2)\right)$. It is clear that the first expression depends on $P(y_1|x)$ and the second expression depends on $P(y_2|x)$ and since for $P(y_1, y_2|x)$ and

$P'(y_1, y_2|x)$ the conditional probability is equal, i.e. $P(y_i|x_i) = P'(y_i|x_i)$, the capacity is the same. ■

C. Capacity region of the Degraded Broadcast channel

Theorem 1 [3, ch. 15.6], [1, ch. 5],[4] The capacity region for sending independent information over the degraded broadcast channel $X - Y_1 - Y_2$ is the convex hull of the closure of all (R_1, R_2) satisfying

$$R_2 \leq I(U; Y_2), \tag{5}$$

$$R_1 \leq I(X; Y_1|U) \tag{6}$$

for some joint distribution $p(u)p(x|u)p(y_1, y_2|x)$, where the additional random variable U has cardinality bounded by $|\mathcal{U}| \leq \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\}$.

We provide a proof fo Theorem 1 for the physically degraded case. However, Theorem 1 holds for the stochastically degraded case as well.

Lemma 2 Theorem 1 holds for the stochastically degraded BC.

Proof: Obviously, it holds for the associated physically degraded version of the BC given by $P'(y_1, y_2|x) = P(y_1|x)P'(y_2|y_1)$ from (3). Here we show that Theorem 1 holds for the probability $P(y_1, y_2|x)$ of the stochastically degraded BC.

The region for the stochastically degraded BC is:

$$R_1 \leq I_{p'}(X; Y_1|U)$$

$$R_2 \leq I_{p'}(U; Y_2).$$

We should show that$I(X; Y_1|U) = I_{p'}(X; Y_1|U)$ and $I(U; Y_2) = I_{p'}(U; Y_2)$

For the phisically degraded BC the distribution is $P(u)P(x|u)P(y_1|x)P(y_2|y_1)$ whereas for the stochastically degraded BC the distribution is $P(u)P(x|u)P(y_1|x)P'(y_2|y_1)$. We denote $a = P(u)P(x|u)P(y_1|x)$ and now we can write the distribution for the physically degraded BC as $a \cdot P(y_2|y_1)$ and the distribution for the stochasticaly degraded BC as $a \cdot P'(y_2|y_1)$ Only the expression a determine the value of $I(X; Y_1|U)$ and $I_{p'}(X; Y_1|U)$ and therefore they are equal. We have that $I(U; Y_2) = I_{p'}(U; Y_2)$ since the expressions depends only on $P(u, y_2)$. We can write $P'(u, y_2) = \sum_x P'(u, x, y_2) = \sum_x P(u)P(x|u)P'(y_2|x)$ and $P(u, y_2) = \sum_x P(u, x, y_2) = \sum_x P(u)P(x|u)P(y_2|x)$. We can chose $P'(y_2|x)$ in a manner that the expressions are equal. ■

We will now find the capacity region of the degraded broadcast channel. Since the proof of the achievability uses unusual tools, we begin the proof from proof of the converse.

proof of convers:

We fix a code $(n, 2^{nR_1}, 2^{nR_2})$ with a probability of error $P_\epsilon^{(n)}$. Now consider,

$$\begin{aligned}
nR_2 &= H(M_2) && (7)\\
&= H(M_2) + H(M_2|Y_2^n) - H(M_2|Y_2^n) && (8)\\
&\overset{(a)}{\leq} I(M_2; Y_2^n) + n\epsilon_n && (9)\\
&= \sum_{i=1}^{n} I(M_2, Y_{2,i}|Y_2^{i-1}) + n\epsilon_n && (10)\\
&\leq \sum_{i=1}^{n} I(M_2, Y_2^{i-1}; Y_{2,i}) + n\epsilon_n && (11)\\
&\overset{(b)}{\leq} \sum_{i=1}^{n} I(M_2, Y_2^{i-1}, Y_1^{i-1}; Y_{2,i}) + n\epsilon_n && (12)\\
&= \sum_{i=1}^{n} I(U_i; Y_{2,i}) + n\epsilon_n. && (13)
\end{aligned}$$

Where

- (a)- Due to Fano inequality $n\epsilon_n \triangleq nRP_\epsilon^{(n)} + 1 \geq H(M_2|Y_2^n)$. Note that since $P_\epsilon^{(n)} \to 0$ then $\epsilon_n \to 0$.
- (b)- Follows from defining $U_i = M_2, Y_2^{i-1}, Y_1^{i-1}$.

Hence we have:

$$\begin{aligned}
R_2 &\leq \frac{1}{n}\sum_{i=1}^{n} I(U_i; Y_{2,i}) + \epsilon_n && (14)\\
&= I(U_Q; Y_2|Q),\ Q \sim U[1, 2, \ldots, n]) && (15)\\
&\leq I(U_Q, Q; Y_2). && (16)
\end{aligned}$$

Now consider,

$$
\begin{aligned}
nR_1 &= H(M_1) && (17)\\
&= H(M_1|M_2) && (18)\\
&= H(M_1|M_2) - H(M_1|M_2, Y_1^n) + H(M_1|M_2, Y_1^n) && (19)\\
&\overset{(a)}{\leq} I(M_1; Y_1^n|M_2) + n\epsilon_n && (20)\\
&= \sum_{i=1}^{n} I(M_1; Y_{1,i}|M_2, Y_1^{i-1}) + n\epsilon_n && (21)\\
&= \sum_{i=1}^{n} I(X_i, M_1, Y_{1,i}|M_2, Y_1^{i-1}) && (22)\\
&\overset{(b)}{=} \sum_{i=1}^{n} I(X_i; Y_{1,i}|M_2, Y_1^{i-1}) && (23)\\
&\overset{(c)}{\leq} \sum_{i=1}^{n} I(X_i; Y_{1,i}|M_2, Y_1^{i-1}, Y_2^{i-1}) && (24)\\
&\overset{(d)}{=} nI(X_1; Y_1|U_Q, Q), && (25)
\end{aligned}
$$

where:

- (a) - By Fano inequality.
- (b) - Because $(Y_{1,i} - X_i - (M_1, M_2, Y_1^{i-1}))$ is a Markov chain.
- (c) - Because $(X_i - (Y_1^{i-1}, M_2) - Y_2^{i-1})$ is a Markov chain (the proof is given at the end of the lecture).
- (d) - Follows the definition of $U_i = M_2, Y_2^{i-1}, Y_1^{i-1}$ and taking $Q \sim U[1, \cdots, n]$.

Remark 1 We notice that $U - X - Y_1 - Y_2$ since $p(y_{1,q}|x_q, u_q, Q = q) = p(y_{1,q}|x_q)$. We also show in Section I-D that $X_i - \left(Y_1^{i-1}, M_2\right) - Y_2^{i-1}$.

■

Remark 2 The cardinality bounds for the random variable U are derived using standard methods from convex set theory, and won't be dealt here.

Remark 3 (Some intuition:) As we can see the theorem above as well as the following proof uses some unusual tools. One of those tools is the auxiliary random variable U.

This random variable functions as a cloud center which can be distinguished by both receivers Y_1 and Y_2. Each cloud consists of 2^{nR_1} codewords X^n distinguishable by Y_1. The worst receiver can see only the clouds, while the better one has the resolution to estimate the specific codeword in the cloud.

Achievability:

Fix $p(u)$ and $p(x|u)$.

Codebook generation: Generate 2^{nR_2} independent codewords of length n, $U(w_2)$, $w_2 \in \{1, 2, \cdots, 2^{nR_2}\}$, according to $\prod_{i=1}^{n} p(u_i)$. For each codeword $U(w_2)$, generate 2^{nR_1} independent codewords $X(w_1, w_2)$ according to $\prod_{i=1}^{n} p(x_i|u_i(w_2))$. Here $u(i)$ plays the role of the cloud center which can be interpreted by both receivers, while $x(i, j)$ is the specific jth codeword in the ith cloud.

Encoding: To send the pair (W_1, W_2), send the corresponding codeword $X(W_1, W_2)$.

Decoding: Receiver 2 determines the unique $\hat{\hat{W}}_2$ such that $(U(\hat{\hat{W}}_2), Y_2) \in A_\epsilon^{(n)}$. If there are none such or more then one such, the receiver declares an error. Receiver 1 looks for the unique $(\hat{\hat{W}}_1, \hat{\hat{W}}_2)$ such that $(U(\hat{\hat{W}}_2), X(\hat{\hat{W}}_1, \hat{\hat{W}}_2), Y_1) \in A_\epsilon^{(n)}$. If there is none such or more then one such, the receiver declares an error.

Error Analysis: By the symmetry of the code generation, the error probability does not depend on the specific codeword which was sent. Hence, we assume that $(W_1, W_2) = (1, 1)$ was sent. We have a single user channel from U to Y_2, we will be able to decode the U codewords with a low probability of error if $R_2 \leq I(U; Y_2)$. We denote $E_2(i) = \{(U^n(i), Y_2^n) \in A_\epsilon^{(n)}\}$. Then the error probability at receiver 2 is

$$P_\epsilon^{(n)}(2) = \Pr\left(E_2^c(1) \bigcup \left(\bigcup_{i \neq 1} E_2(i)\right)\right) \tag{26}$$

$$\leq P(E_2^c(1)) + \sum_{i \neq 1} P(E_2(i)) \tag{27}$$

$$\overset{(a)}{\leq} \epsilon + 2^{nR_2} 2^{-nI(U;Y_2) - 2\epsilon} \tag{28}$$

$$\leq 2\epsilon \tag{29}$$

if n is large enough and $R_2 \leq I(U; Y_2)$, where (a) follows from AEP. Similarly, for receiver 1, we define

$$\begin{aligned} E_1(i) &= \{(U^n(i), Y_1^n) \in A_\epsilon^{(n)}\}, &(30)\\ E_1(i,j) &= \{(U^n(i), X^n(i,j), Y_1^n) \in A_\epsilon^{(n)}\}. &(31)\end{aligned}$$

Recall that receiver 2 determines the uniqe $\hat{\hat{W}}_2$ such that $(U(\hat{\hat{W}}_2), Y_2) \in A_\epsilon^{(n)}$ and then receiver 1 looks for the unique $(\hat{\hat{W}}_1, \hat{\hat{W}}_2)$ such that $(U(\hat{\hat{W}}_2), X(\hat{\hat{W}}_1, \hat{\hat{W}}_2), Y_1) \in A_\epsilon^{(n)}$. Hence, we first established that $P_\epsilon^{(n)}(2) \to 0$. We can bound the probability of error as

$$\begin{aligned} P_\epsilon^{(n)}(1) &= P\left(E_1^c(1) \bigcup E_1^c(1,1) \bigcup \bigcup_{i\neq 1} E_1(i) \bigcup \bigcup_{j\neq 1} E_1(1,j)\right) &(32)\\ &\leq P(E_1^c(1)) + P(E_1^c(1,1)) + \sum_{i\neq 1} P(E_1(i)) + \sum_{j\neq 1} P(E_1(1,j)). &(33)\end{aligned}$$

By the same arguments as for receiver 2, we can bound $P(\tilde{E}_{Yi}) \leq 2^{-n(I(U;Y_1)-3\epsilon)}$. So, the term $\sum_{i=2}^{2^{nR}} P(E_i)$ goes to 0 if $R_2 < I(U;Y_1)$. But, by the data-processing inequality and due to the fact the it is degraded channel, $I(U;Y_1) \geq I(U;Y_2)$, and hence the conditions of the theorem imply that the term $\sum_{i\neq 1} P(E_i)$ goes to 0. We can also bound the fourth term as

$$\begin{aligned} P(E_{1j}) &= P((u^n(1), x^n(1,j), y_1^n) \in A_\epsilon^{(n)}) &(34)\\ &= \sum_{(U,X,Y_1)\in A_\epsilon^{(n)}} P(u^n(1))P(x^n(1,j)|u^n(1))P(y_1^n|u^n(1)) &(35)\\ &\leq \sum_{(U,X,Y_1)\in A_\epsilon^{(n)}} 2^{-n(H(U)-\epsilon)}2^{-n(H(X|U)-\epsilon)}2^{-n(H(Y_1|U)-\epsilon)} &(36)\\ &\leq 2^{n(H(U,X,Y_1)+\epsilon)}2^{-n(H(U)-\epsilon)}2^{-n(H(X|U)-\epsilon)}2^{-n(H(Y_1|U)-\epsilon)} &(37)\\ &= 2^{-n(I(X;Y_1|U)-4\epsilon)}. &(38)\end{aligned}$$

Hence if $R_1 < I(X;Y_1|U)$, the fourth term goes to 0. Thus, the probability of error is bounded by

$$P_\epsilon^{(n)}(1) \leq \epsilon + \epsilon + 2^{nR_2}2^{-n(I(U;Y_1)-3\epsilon)} + 2^{nR_1}2^{-n(I(X;Y_1|U)-4\epsilon)} \quad (39)$$

$$\leq 4\epsilon \quad (40)$$

if n is large enough and $R_2 < I(U;Y_2)$ and $R_1 < (X;Y_1|U)$.

Alternatively, decoder 1 can work in two stages. The first stage is to decode W_2 by finding $\hat{W}_2$ such that $(U^n(\hat{W}_2), Y_1^n) \in A_\epsilon^{(n)}$. The second stage, after decoding W_2, is to decode W_1 by finding $\hat{W}_1$ such that $(U^n(\hat{W}_2), X^n(U^n(\hat{W}_2), \hat{W}_1), Y_1^n) \in A_\epsilon^{(n)}$. We can bound the probability of error using the following observation: $\Pr(A \cup B) \leq \Pr(A) + \Pr(B|A^c)$. Note that if $\Pr(A^c) = 0$ it follows immediately. Else, we have that

$$\begin{aligned} \Pr(A \cup B) &= \Pr(A) + \Pr(B \cap A^c) \\ &\leq \Pr(A) + \frac{\Pr(B \cap A^c)}{\Pr(A^c)} \\ &= \Pr(A) + \Pr(B|A^c). \end{aligned}$$

If we take the event A to be the event $\{E_1^c(1) \bigcup \bigcup_{i\neq 1} E_1(i)\}$ and the event B to be the event $\{E_1^c(1,1) \bigcup \bigcup_{j\neq 1} E_1(1,j)\}$ we have that $\Pr(A) \to 0$ when $R_2 \leq I(U;Y_2)$ and $\Pr(B|A^c) \to 0$ when $R_1 \leq I(X;Y_1|U)$. Using the mentiond observation we have that $\Pr(A \cup B) \to 0$.

The above bounds show that we can decode the messages with total probability of error that goes to 0. With this, we complete the proof of the achievability of the capacity region for the degraded broadcast channel. ■

D. sufficient condition for verifying Markov chains

This method was first introduced in [2]. Assume a set of random variables $(X_1, X_2, \cdots, X_N)$. Without loss of generality we assume that the joint distribution has the form

$$p(x^N) = f(x_{\mathcal{S}_1}) f(x_{\mathcal{S}_2}) \cdots f(x_{\mathcal{S}_k}). \quad (41)$$

Where $\mathcal{S}_i$ is a subset of $\{1, 2, \cdots, N\}$ and $X_{\mathcal{S}_i} = \{X_j\}_{j \in \mathcal{S}_i}$. The following graphical technique provides a sufficient condition for Markov relation $X_{\mathcal{G}_1} - X_{\mathcal{G}_2} - X_{\mathcal{G}_3}$ where $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$ are disjoint subsets of X^N. The technique comprises two steps:

- Draw an undirected graph where all the random variables $X_i, i \in \{1, 2, \cdots, N\}$ are nodes in the graph and for all $i \in \{1, 2, \cdots, k\}$ draw edges between all the nodes $X_{\mathcal{S}_i}$.
- If all paths in the graph from a node in $X_{\mathcal{G}_1}$ to a node in $X_{\mathcal{G}_3}$ pass throw a node in X_{i2} then the Markov chain $X_{\mathcal{G}_1} - X_{\mathcal{G}_2} - X_{\mathcal{G}_3}$ holds.

We will demonstrate it by an example.

Example 1 (Markov chain) Assume

$$p(x^2, y^2, z^2) = p(x_1, y_2)p(y_1, x_2)p(z_1|x_1, x_2)p(z_2|y_1), \tag{42}$$

is $X_1 - X_2 - Z_2$ Markov?
To verify that it is Markov, we draw all the connections between x_i, y_i, z_i according to the functions. For every function, we draw a line on the function's arguments. if there is a line connecting $x_1 - y_1 - z_2$ and not directly $x_1 - z_2$ then it is Markov. In this example we have such a line so it is Markov.

Now we show that $X_i - \left(Y_1^{i-1}, M_2\right) - Y_2^{i-1}$. Consider,

$$p(m_1, m_2, x^i, y_1^{i-1}, y_2^{i-1}) = p(m_1)p(m_2)p(x^{i-1}|m_1, m_2)p(x_i|m_1, m_2)p(y_1^{i-1}|x^{i-1})p(y_2^{i-1}|y_1^{i-1}). \tag{43}$$

Following the steps metioned earlier we obtain the scheme in Fig. 3 and by using the explanations above we can derive that $X_i - \left(Y_1^{i-1}, M_2\right) - Y_2^{i-1}$. This gives us the wanted result.

REFERENCES

[1] A. E. Gamal and Y. H. Kim, *'Lecture Notes on Network Information Theory'* . http://arxiv.org/abs/1001.3404

[2] H. H. Permuter and Y. Steinberg and T. Weissman, *'Two-Way Source Coding With a Helper'*. IEEE Transactions on Information Theory, 2010, June, Vol. 56.

[3] T. M. Cover and J. A. Thomas, *'Elements of Information Theory.'*. Wiley, New York, 2nd edition 2006

[4] T. M. Cover, *'Broadcast channels'*. IEEE Transactions on Information Theory, 1972, Jan, Vol. 18.

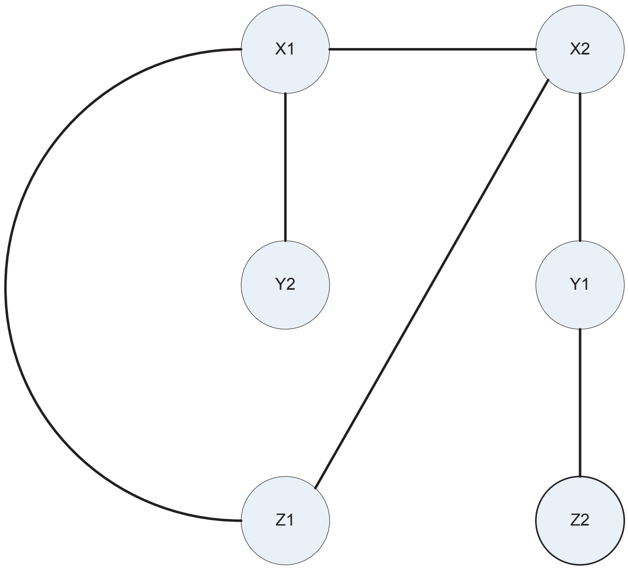

Fig. 2. Diagram that represent the Markov connections of Example 1.

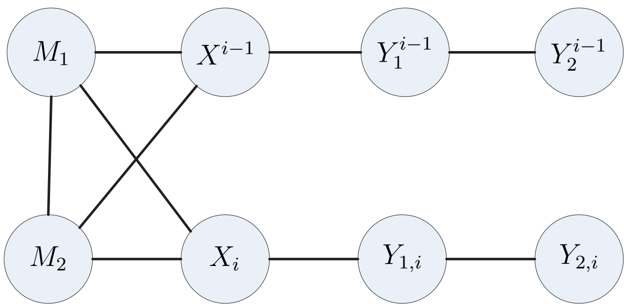

Fig. 3. Markov connections of $p(m_1)p(m_2)p(x^{i-1}|m_1,m_2)p(x_i|m_1,m_2)p(y_1^{i-1}|x^{i-1})$ from which we can derive $X_i - \left(Y_1^{i-1}, M_2\right) - Y_2^{i-1}$ using the steps and explanations mentioned above.